MINI PROJECT

FITTING OF DISTRIBUTION

December 27, 2020

Benitta Susan Aniyan
Enrollment Number: EBEON1020323564
Batch Number: 2020-4868

DATA DESCRIPTIVE: Secondary data of 20 observations of gold price history in India during the period 1950 to 1970 from web source (reference: http://welcomenri.com/gold/gold50years-history.aspx)

YEAR	1950	1951	1952	1953	1954	1955	1956	1957	1958
RATE	99	98	76	73	77	79	90	90	95

YEAR	1959	1960	1961	1962	1963	1964	1965	1966	1967
RATE	102	111	119	119	97	63	71	83	102

YEAR	1968	1969	1970
RATE	162	176	184

OBJECTIVE OF THE STUDY:
To find the best fit distribution for the given dataset.

CODE:
The analysis is done by using R software.

The codes are given below:

```
library('fitdistrplus')
library('logspline')
library('actuar')
data=read.csv(file="gold data.csv", header=TRUE)
summary(data)
plotdist(data$rate, histo = TRUE, demp = TRUE)
descdist(data$rate, boot=1000)
boxplot(data$rate)
weibull=fitdist(data$rate, "weibull")
summary(weibull)
normal=fitdist(data$rate, "norm")
summary(normal)
lognormal=fitdist(data$rate, "lnorm")
summary(lognormal)
gamma=fitdist(data$rate, "gamma")
summary(gamma)
logistic=fitdist(data$rate, "logis")
summary(logistic)
pareto=fitdist(data$rate, "pareto",, start = list(shape = 0.3, scale = 500))
summary(pareto)
burr=fitdist(data$rate, "burr", start = list(shape1 = 0.3, shape2 = 1,rate = 1))
summary(burr)
```

```
cdfcomp(list(weibull, normal, lognormal, gamma, logistic, pareto, burr), xlogscale = TRUE, ylogscale = TRUE, legendtext = c("weibull", "normal", "lognormal", "gamma", "logistic", "pareto", "burr"))
gofstat(list(weibull, normal, lognormal, gamma, logistic, pareto, burr), fitnames = c("weibull","normal", "lognormal", "gamma", "logistic","pareto", "burr"))
plot.legend <- c("weibull","normal", "lognormal", "gamma", "logistic","pareto", "burr")
denscomp(list(weibull,normal,lognormal,gamma,logistic, pareto, burr), legendtext = plot.legend)
```

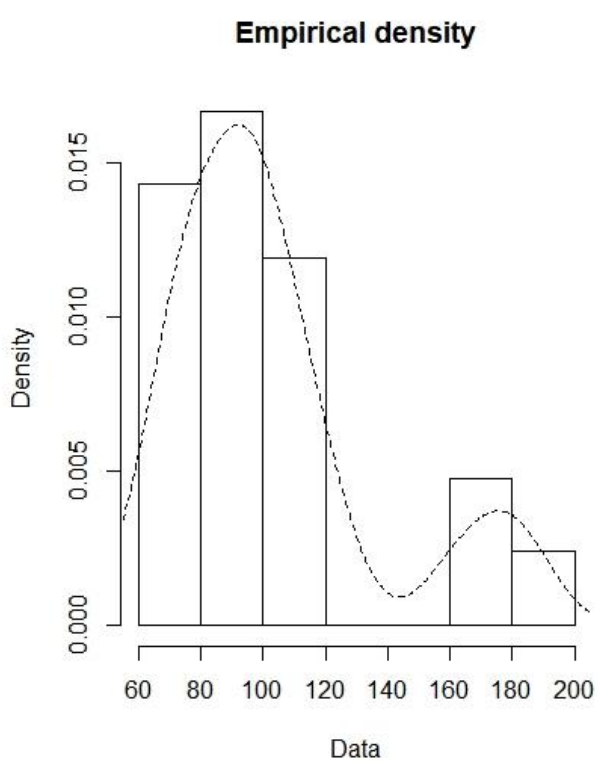
PROCEDURE:

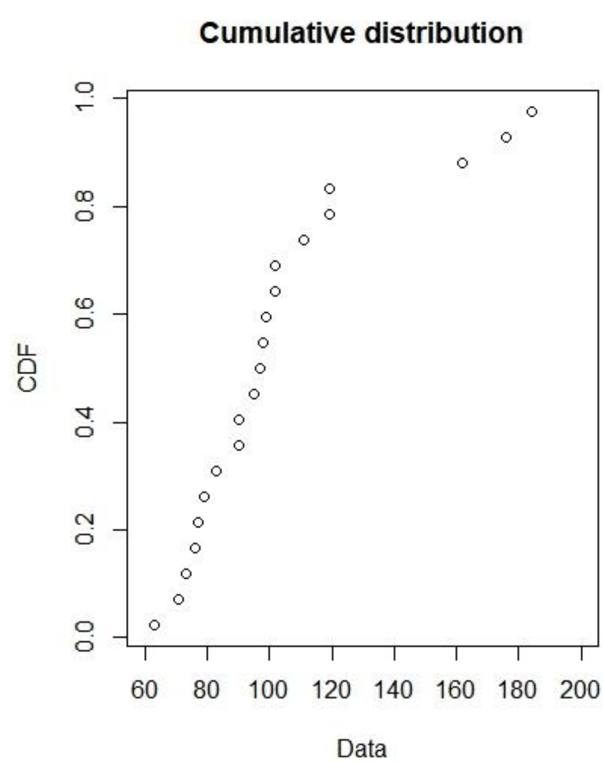
The basic summary of the data is given in the following table.

Min	1^{st} Quartile	Median	Mean	3^{rd} Quartile	Max
63.0	79.0	97.0	103.1	111.0	184.0

Here the mean is near to 3^{rd} quartile so it is approximately positively skewed so it may be like **Weibull, Gamma, Logistic, Lognormal**, **Pareto**, **Burr** distributions.

Density and CDF:





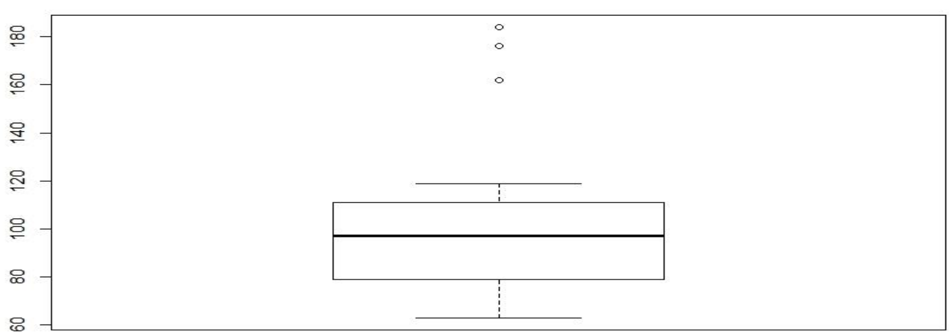
Descriptive Statistics

Statistic	Value
Sample Size	21
Range	121
Mean	103.14
Variance	1116.6
Std. Deviation	33.416
Coeff. of Variation	0.32398
Std. Error	7.292
Skewness	1.3875
Excess Kurtosis	1.3191

Percentile	Value
Min	63
5%	63.8
10%	71.4
25% (Q1)	78
50% (Median)	97
75% (Q3)	115
90%	173.2
95%	183.2
Max	184

Here the coefficient of skewness>0, so the data is positively skewed.

BOX PLOT:



From the above box plot, most of the observations are lie above the median, so it is positively skewed.

Fitting of the Weibull distribution by maximum likelihood
Parameters :

	Estimate	Standard Error
Shape	3.210803	0.4981175
Scale	114.947308	8.3196257
Loglikelihood:- -103.1507	AIC:-210.3015	BIC:-212.3905

Fitting of the Normal distribution by maximum likelihood
Parameters :

	Estimate	Standard Error
mean	103.14286	7.116232
Standard deviation	32.61067	5.031936
Loglikelihood:- -102.9751	AIC:-209.9503	BIC:-212.0393

Fitting of the Lognormal distribution by maximum likelihood
Parameters :

	Estimate	Standard Error
meanlog	4.5932445	0.06178260
Standard deviation log	0.2831234	0.04368444
Loglikelihood:- -99.75653	AIC:-203.5131	BIC:-205.6021

Fitting of the Gamma distribution by maximum likelihood
Parameters :

	Estimate	Standard Error
shape	11.8208681	3.59350506
rate	0.1146141	0.03558916
Loglikelihood:- -100.6123	AIC:-205.2247	BIC:-207.3137

Fitting of the Logistic distribution by maximum likelihood
Parameters :

	Estimate	Standard Error
location	97.91863	6.280570
scale	16.95166	3.223181
Loglikelihood:- -102.1717	AIC:-208.3434	BIC:-210.4324

Fitting of the Pareto distribution by maximum likelihood

Parameters :

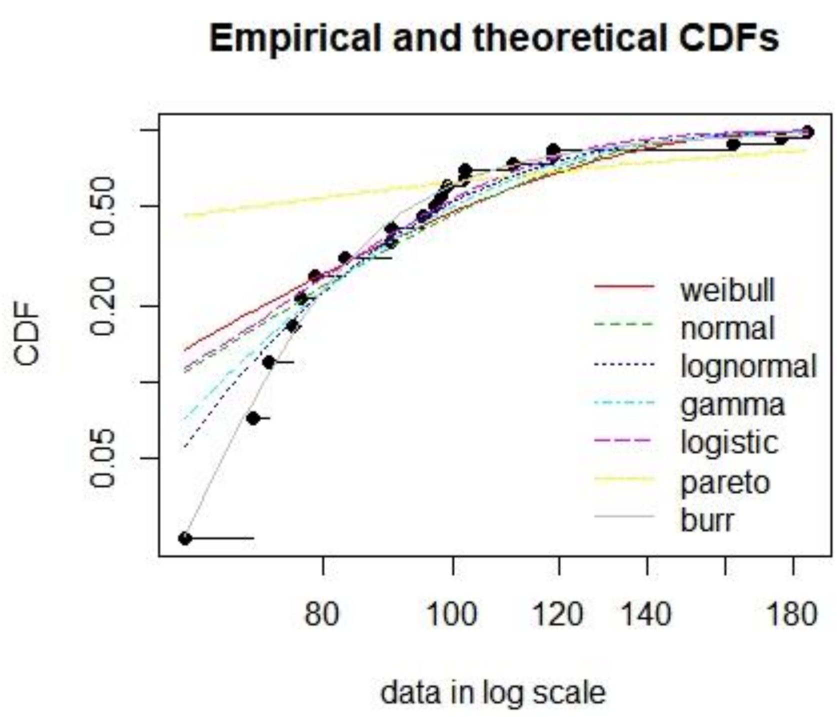
	Estimate	Standard Error
shape	11043775	N/A
scale	1139018201	N/A
Loglikelihood:- -118.3584	AIC:-240.7168	BIC:-242.8059

Fitting of the Burr distribution by maximum likelihood

Parameters :

	Estimate	Standard Error
Shape1	0.32516101	0.256039121
Shape2	11.54002413	6.033197046
Loglikelihood:- -98.57228	AIC:-203.1446	BIC:-206.2781

The Empirical and Theoretical CDF of distributions given below:



From the graph it is clear that Burr distribution is more close to the observations.

Goodness-of-fit statistics

	WEIBULL	NORMAL	LOGNORMAL	GAMMA	LOGISTIC	PARETO	BURR
KOLMOGOROV-SMIRNOV STATISTIC	0.2202244	0.228264	0.1696715	0.190677	0.15438354	0.4571053	0.10288493
CRAMER-VON MISES STATISTIC	0.2060290	0.2013785	0.0824739	0.1158826	0.082687571	1.0856767	0.03530912
ANDERSON-DARLING STATISTIC	1.2404035	1.2645822	0.5840735	0.7808309	0.87358151	5.1713530	0.22804662

From the above table it is clear that the Kolmogorov-Smirnov, Cramer-Von Mises and Anderson-Darling Statistic for Burr distribution is less than the other distributions. So, Burr distribution is a good fit for the data.

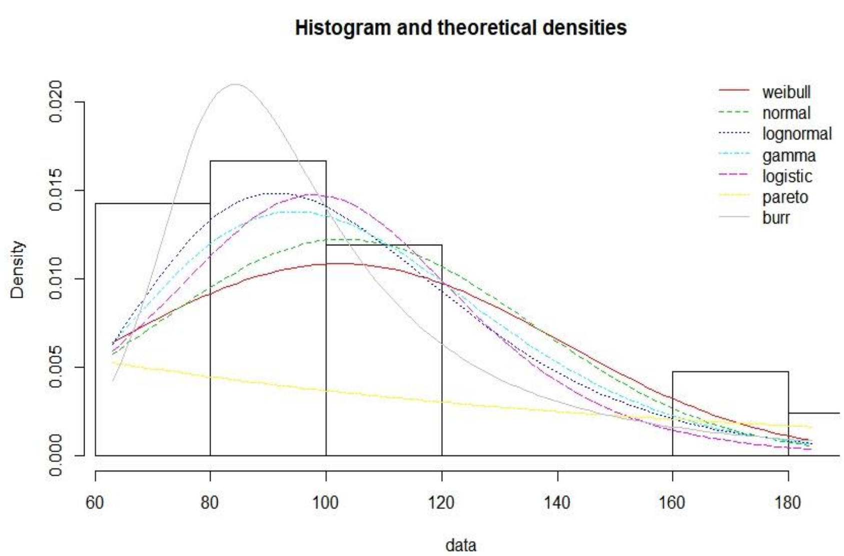
Goodness-of-fit criteria

	WEIBULL	NORMAL	LOGNORMAL	GAMMA	LOGISTIC	PARETO	BURR
AIC	210.3015	209.9503	203.5131	205.2247	208.3434	240.7168	203.1446
BIC	212.3905	212.0393	205.6021	207.3137	210.4324	242.8059	207.2781

Where, AIC is Akaike's Information Criterion and BIC is Bayesian Information Criterion

While comparing with these distributions, since the AIC and BIC of the Burr distribution are minimum. So the approximate fitted distribution will be Burr distribution.

DENSITY COMPARISON:



Here, the graph of Burr distribution is approximately normal. So, the Burr distribution is a best fit for the data.

CONCLUSION:-

From the Goodness-of-fit statistics and criteria, we conclude that the Burr distribution is the best fit for the gold price data from 1950 to 1970.

The estimated parameters of the data for the Burr distribution are:

	Estimate	Standard Error
Shape1	0.32516101	0.256039121
Shape2	11.54002413	6.033197046
Loglikelihood:- -98.57228	AIC:-203.1446	BIC:-206.2781

Goodness-of-fit- statistics

	Test statistic value
KOLMOGOROV-SMIRNOV STATISTIC	0.10288493
CRAMER-VON MISES STATISTIC	0.03530912
ANDERSON-DARLING STATISTIC	0.22804662

Goodness-of-fit criteria

	BURR
AIC	203.1446
BIC	207.2781